(1) Publication number:

0 079 739

A2

12

EUROPEAN PATENT APPLICATION

(21) Application number: 82305926.6

(22) Date of filing: 08.11.82

(5) Int. Cl.³: C 12 N 15/00

C 12 N 1/00, C 12 P 21/02 C 07 H 21/04, C 07 C 103/52 //C12R1/19, C12R1/865

(30) Priority: 12.11.81 US 320632

Date of publication of application: 25.05.83 Bulletin 83/21

Designated Contracting States: BE CH DE FR GB IT LI NL SE (1) Applicant: THE UPJOHN COMPANY 301 Henrietta Street Kalamazoo, Michigan 49001(US)

(72) Inventor: Dugalczyk, Achilles e/o The Upjohn Company 301 Henrietta Street Kalamazoo Michigan 49001(US)

(4) Representative: Perry, Robert Edward et al, GILL JENNINGS & EVERY 53-64 Chancery Lane London WC2A 1HN(GB)

(S) Albumin-based nucleotides, their replication and use, and plasmids for use therein.

(5) The DNA sequence coding for human serum albumin has been isolated and inserted as two fragments into two novel plasmids which can be replicated in *E. coli*. These novel fragments can be joined to provide a unitary DNA sequence which then can be cloned into a suitable host, e.g. *E. coli*, for the expression of human serum albumin (which is used extensively in medical practice in treating shock conditions).

EP 0 079 739 A

UPJOHN

\$

1

ALBUMIN-BASED NUCLEOTIDES, THEIR REPLICATION AND USE, AND PLASMIDS FOR USE THEREIN

This invention relates to nucleotides related to human serum albumin (HSA), their replication and use, and plasmids (and host substances) for use therein.

The gene for serum albumin is regulated in

5 development. On the other hand, serum albumin is synthesised in mammals by the adult liver, and its plateau in adulthood. The embryonic liver and yolk sac, on the other hand, produce predominantly a-fetoprotein, but the synthesis decreases drastically after birth. Recently,

10 Law et al determined the complete sequence of mouse

α-fetoprotein mRNA, Nature 291 (1981) 201-205. The structure revealed extensive homology to mammalian serum albumin, indicating that the two proteins are encoded in the same gene family. Similar conclusions have been

15 reached from studies on the α -fetoprotein genes of the rat and the mouse; see Jagodzinski et al, Proc. Natl. Acad. Sci. USA, 78 (1981) 3521-3525, and Gorin et al, J. Biol. Chem. 256 (1981) 1954-1959.

The complete nucleotide sequence of human serum

20 mRNA has been determined from recombinant cDNA clones and
from a primer-extended cDNA synthesis on the mRNA
template. The sequence comprises 2,078 nucleotides,
starting upstream of a potential ribosome binding site
in the 5'-untranslated region. It contains all the

- 25 translated codons and extends into the poly(A) at the 3'-terminus. Part of the translated sequence codes for a hydrophobic prepeptide met-lys-trp-val-thr-phe-ile-ser-leu-leu-phe-leu-phe-ser-ser-ala-tyr-ser, followed by a basic propeptide arg-gly-val-phe-arg-arg. These signal
- 30 peptides are absent from mature serum albumin and, so far, have not been identified in their nascent state in humans. A remaining 1,755 nucleotides of the translated mRNA sequence code for 585 amino acids which are in agreement, with few exceptions, with the published amino
- 35 acid data for human serum albumin. The mRNA sequence verifies and refines the repeating homology in the triple-domain structure of the serum albumin molecule.

DETAILED DESCRIPTION OF THE INVENTION

Human serum albumin cDNA is cloned into the PstI site of plasmid pBR322 by the oligo(dG)-oligo(dC) tailing technique. Plasmid DNA was isolated from 97 positive colonies which hybridized to the enriched albumin cDNA probe, and the recombinant plasmid pHA36 was found to contain the largest insert of an albumin cDNA sequence. Its restriction endonuclease map is shown in the drawing, together with a restriction map of the primer-extended plasmid clone pHA206. The latter was obtained in a second transformation experiment after initiating the cDNA synthesis from an internal primer. This primer was a 91 base pairs long DNA fragment, MspI(152)-TaqI(182/3), isolated from pHA36. The two plasmids, pHA36 and pHA206, share 0.15 kb of homologous DNA. Together, they encode the entire sequence for human serum albumin, starting with the CTT codon for leu -10 of the prepeptide and extending into the 3'-untranslated region of poly(A).

Sequence of the Albumin cDNA. The sequence was determined for the most part on both DNA strands to ensure accuracy. All of the restriction sites used to end-label DNA fragments were sequenced across by 20 labeling a neighboring restriction site. The entire nucleotide sequence of the serum albumin mRNA, as determined from the cloned DNA in pHA36, pHA206, and from the primer-extended cDNA at the 5'-terminus of the message, is shown in the following Table 1. The inferred amino acid sequence is also indicated. The mRNA length is 2,078 nucleo- 25 tides, of which 38 represent the 5'-untranslated region, 54 identify a prepeptide of 18 amino acids, 18 identify a propeptide of 6 amino acids, 1,755 code for the known 585 amino acids of serum albumin, 189 make up the 3'-untranslated region and 24 are the poly(A) sequence. Nucleotides 5 to 15 (-34 to -24) in the 5'-untranslated region (Table 1) are complementary to a 3'-terminal region of eukaryotic 18S RNA [Azad, A.A. and Deacon, N.J. (1980) Nucl. Acids Res. 8, 4365-4376] and thus could represent a ribosome binding site:

(5')...T
$$T^CT$$
 C T T C T G T......albumin mRNA (3')...G A G G A A G G C G U C C m_2^6 A m_2^6 A.....18S RNA

35

The translated portion of the mRNA sequence codes for the signal peptide and the main body of the albumin polypeptide chain. The

signal peptide is composed of a hydrophobic prepeptide of 18 amino acids and a basic propeptide of 6 amino acids (Table 1). Since prepeptides are removed from nascent secretory proteins (like albumin) in the endoplasmic reticulum, they are seen only in vitro in heterologous translation systems. As yet, they have not been found within cells [Judah, J.D. and Quinn, P.S. (1977) FEBS 11th Mtg., Copenhagen 50, 21-29; and Strauss, A.W., Donohue, A.M., Bennett, C.D., Rodkey, J.A. and Alberts, A.W. (1977) Proc. Natl. Acad. Sci. USA 74, 1358-1362]. This is the first report of the presence and the sequence of a prepeptide for human serum albumin. As it is with other secretory proteins, the conversion of proalbumin to albumin takes place in the Golgi vesicles, and the enzyme responsible for this cleavage is probably cathepsin B [Judah, J.D. and Quinn, P.S. (1978) Nature 271, 384-385]. This is also a first report on the sequence of the propeptide for normal human serum albumin.

At the 3'-end of the message, the putative polyadenylation signal sequence, AATAAA, is located 164 nucleotides downstream from the amino acid termination codon TAA and 16 nucleotides upstream from the beginning of the poly(A) sequence. Another characteristic sequence located near the polyadenylation site has been identified by Renoist, et al. [Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl. Acids Res. 8, 127-142]; the concensus sequence from several mRNAs was concluded as TTTTCACTGC. A similar sequence, TTTTCTCTGT, is located 19 nucleotides upstream from the AATAAA hexanucleotide in the human albumin mRNA (Table 1).

TABLE 1

	(30)	(170)	260)	(350)	(440)	(330)	(029)	(710)	(300)
_	AGC	20 173 AA (50 ala GCA (260)	1eu C	110 pro ccA (140 try TAT (170 gln CAA (200 cys TCT (230 alu GAA (
5	phe s TTT A	phe 1 TTC A		thr 1 ACT C	asn p	leu t	169 1 cys q TGC C		ela GCA C
	leu p	asn p AAT T	alu phe CAA TTT	ala t GCA A	e ase GAC A	tyr 1 TAC 1	168 1 cys c 1CT 1	leu lys CTC AAG	phe a
	phe 1 TTT C	9.1° 6. ★2			asp a		-		alu p GAG 1
	<u>a</u> ⊢	9 16 GA G	val thr GTA ACT	thr v	Jys a	2 X	thr glu ACA GAA	qin arq CAG AGA	ele qlu CCT CAC
10	-10 Jeu Jeu phe CTT CTT TTT	aly a	glu val thr GAA GTA ACT	75 gly asp lys leu cys thr val CGA GAC AAA TTA TGC ACA GTT	his J CAC A	leu lys lys TTG AAA AAA	he t	1 × 8	175 AA 0
20	ser) TCC C	ភិ ភិ ភ	A G	eu e	بر چ او چ	phe 1 TTT 1	la p	ala l	2 S
		Sp 1	al a	X 1 X 1	5 5 5	thr phe ACA TTT	ala ala phe GCT GCT TTT	ser ala lys TCT GCC AAA	phe pro TTT CCC
	1 k	* \$ \$ \$	3 ¥ 7	asp lys leu GAC AAA TTA	he 1 TC T	glu t	lys a	ser s	5.43
	p r o trp val tlu phe lle ser TGG GTA ACC TTT ATT TCC	phe lys asp leu TTT AAA GAT TTG	40 val lys leu val asn GTA AAA TTA GTG AAT	41,000	100 101 glu cys phe leu gln hls GAA TCC TTC TTG CAA CAC	glu glu thr GAA GAG ACA	160 arg tyr lys ala ala phe AGG TAT AAA GCT GCT TTT	SCT 1	gin arg phe pro lys CAG AGA TTT CCC AAA
15	p r val t GTA A	10 arg p ccc 1	6 5 1 1 7 7	5 # E	100 1 CAA 1	130 asn g	160 arg t AGG 1	190 1ys a	220 Ser 9
	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	ls a	Mas v	70 Jeu phe CTT TTT	asn c	S S S S S S S S S S S S S S S S S S S	% X X	7, 80	220 leu ser CTG AGC
		ala his GCT CAT	40 asp his val lys leu val asn GAT CAT GTA AAA TTA GTG AAT	rhr)		2 1 S	ala lys CCT AAA	afe 8	2 2
	-18 Met lys ATG AAG		28	62 cys asp lys ser leu his thr leu phe TGT GAC AAA TCA CTT CAT ACC CTT TTT	aly arq ccc AcA	phe 1	phe TTT	190 asp qlu qly lys ala GAT GAA GGG AAG GCT	220 lys ale trp ale val ale arg leu ser AAA GCA TGG GCA GTA GCT CGC CTG AGC
20		glu val GAG GTT	phe T	2 E		ala CCT	phe TTC	500	val CTA
20	7665	ala his lys ser glu GCA CAC AAG AGT GAG	pro phe CCA TTT	asp lys ser leu GAC AAA TCA CTT	glu pro CAA CCT	Act th	glu leu leu phe GAA CTC CTT TTC	GT E	lys ala trp ala val AAA GCA TGG GCA GTA
	cctt	N/s	34 033 TGT	<u>\$</u>	£ 8	124 938 1GC	leu CTC	g S	5 5
	Ş	ala hís lys GCA CAC AAG	34 gin gin cys CAG CAG TGT	asp	1ys	met ATG	₹ 8	asp	8 5 CCA
	V	ala CCA	gln	62 cys TGT	ala lys CCA AAA	val CTC	ord CCG	5 CTC	\$ ₹
25	GCTITICTCTTCTGTCAACCCCACAGCCCTTTGGCACA	asp GAT			91 cys TGT	asp val CAT GTG	ala CCC	180 pro lys leu asp glu leu arg CCA AAG CTC GAT GAA CTT CGG	
	стст	-1 1 arg asp CGA GAT	30 tyr leu TAT CTT	60 glu asn GAA AAT	90 cys TGC	120 val GTT	150 tyr TAT	180 CCA	210 glu arg ala phe GAA AGA GCT TTC
	TCTT	arg CGT	gla CAG	ala GCT	asp GAC	ale GAG	phe TT	leu leu CTG TTG	gly glu arg a
	1110	phe TTT	ala CCT	<u>ام</u> ع	ala CCT	5 X	Ϋ́ς	leu CTG	ag ¥5
	52	val GTG	phe	g) to	aet ATG	årg AS	pro CCT	177 978 160	91¢
30		oly cct	နှုန် တွင	asp	£ §	leu val TTG GTG	arg his pro t	al a ငင်	ag E
		9- 9- 9-000		ala GCT	aly CCT	1eu 11G	5 S	ala CCT	lys AAA
		 ser 700	15		tyr IAT	år9 CGA	9 3 3 4 5 4 5 4 5 4 5 4 5 4 5 4 5 4 5 4 5	175	ag VS
		tyr IAT	21 ala leu val leu lle GCC TIG GTG TTG ATT	51 53 lys thr cys val AAA ACA TGT GTT	th ACC	org CCC	ile ala arg ATT GCC AGA	ala asp lys ala CCT CAT AAA CCT	201 ala ser leu gin lys phe GCC AGT CTC CAA AAA TTT
35		ala GCT		당		leu CTC	11e ATT		ser
		3er 100	21 ala GCC	22 ¥3 &	81 arg CCT	111 asn AAC	₹ 5 %	171 ala CCT	201 ala GCC

	(890)	(980)	320 ala CCT (1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
	260 leu CTT	290 11e ATT	320 ala GrT	350 818 CCC	380 leu CCT	410 arg CGT	440 hts	&70 ser AGT	500 173 AAA	530 val CTT
5		289 cys 107	tyr		pro CCT	val GTT			val pro GTT CCC	CTT
	ala asp asp arq ala asp GCT GAT GAC AGG GCG GAC	hts	asn tyr AAC TAT	leu arg leu CTG AGA CTT	1 ys	ie i	438 cys lys TGT AAA	pro val CCA GTA	val	ala leu CCA CTT
	ard AGG		1ys AAA	leu CTG		Jeu CTG	437 cys TGT	thr	tyr TAC	
	asp arq GAC AGG	lys ser AAA TCC	316 cys TG ^r	val leu leu leu GTG CTG CTG	alu phe GAA TTT	ala RC	133	460 461 leu ser val val leu asn gln leu cys val leu his glu lys thr CTA ICC GTG GTC CTG AAC CAG TTA TGT GTG TTG CAT GAG AAA ACG	asp glu thr tyr GAT GAA ACA TAC	lys lys ain thr AG AA CA ACT
10	ala asp CCT CAT	leu glu TTG CAA	val GTT	Jeu CTG	asp	qin asn CAG AAT	aly ser GGC AGC	hís glu CAT GAG	glu thr GAA ACA	lys AA
	ele CCT	Jeu TTG	asp CAT	val GTG	phe TTC	aşu CYC	41 <i>y</i>	Z Z	asp	lys AG
	253 cys TGT	leu CTG	glu ser lys GAA AGT AAG	val GTC	val	phe TTC	val	leu TTG	val	ain ile CAA ATC
	2 ₹	pro	ser	36r TCT	2× ₹	tyr lys TAC AAA	8 }	val GTG	ale GA	E &
•	250 253 asp leu leu glu cys GAT CTG CTT GAA TGT	279 280 cys qlu lys fGT GAA AAA		tyr TAC	369 370 ala ala asp pro his qiu cys tyr ala lys val GCT GCA GAT CCT CAT GAA TGC TAT GCC AAA GTG	400 glu tyr lys GAG TAC AAA	430 leu qly CTA GGA	461 cys TGT	490 ala leu GCT CTG	S20 ser qlu lys qlu arq qln lle TCT GAG AAG GAG AGA CAA ATC
15	250 asp leu leu GAT CTG CTT	279 280 cys qlu lys TGT GAA AAA	310 phe val TTT CTT	340 esp tyr GAT TAC	369 370 cys tyr IGC 1AT	400 aly alu GGA GAG	430 1eu CTA	460 461 gln leu cys CAG TTA TGT	490 ala CCT	520 4 lu GAG
		279 cys 1GT	ag t	pro CCT	369 cys TGC	÷ 55	AC	95 CAG	phe ser TTT TCA	1ys AAG
	his gly CAT GGA	278 035 TGC	es TA	arg his AGG CAT	£ 8	phe alu aln leu TTT CAG CAG CTT	arg AGA	AAC	cys phe TGC TTT	5 S
	CAT S	3	ala	arg AGC	\$ 5	44	16. 16.	Je CTG	cys TGC	Ser TCT
20	245 246 thr glu cys cys ACG GAA TGC TGC	leu lys CTG AAG	ala CCT	ala arq GCA AGA	pro CCT	9 of 0	glu val GAC GTC	val	bro CCA	CT C
20	245 cys TOC		Je TY		asp			val	క్తి న్ర	thr ACA
	ole SA	ser lys AGT AAA	pro ser CCT TCA	glu tyr GAA TAT	ala GCA	glu leu GAG CTT	leu val CTT GTA	196 107	asn arg AAC AGG	\$ \$ 55 5 55
	thr Acc				ala GCT			leu CTA		11e ATA
	h is	367 100	leu TTG	tyr	ala occ	392 cys TGT	thr	1	leu val TTG GTG	asp CAT
25	val GTC	11e ATC	asp GAC	leu TTG	361 cys TGT	390 gln asn CA AAT	org CCA	asp CAC	leu TTG	<u>۾</u> کي
	240 thr lys val ACC AAA GTC	270 3er TCG	300 ala ccT	330 phe TTT	360 cys TGC		420 thr ACT	450 glu CA	480 ser TCC	510 h1s CAT
	val thr asp leu thr GTG ACA GAT CTT ACC	asp CAT	pro CCT	met ATG	1ys AAG	1 × 3	ser TCA	ala GCA	91c	9 kg
	leu CTT	gla CA	glu met GAG ATG	leu gly TTG GGC	91 ⁰	11e ATC	val GTG	448 0ys TGT	thr ACA	ACC
	asp CAT	asn AAT	ale GAG	leu 11G	leu CTA	gin asn leu ile CAG AAT TTA ATC	g¥ S	448 met pro cys ATG CCC TGT	476 477 1ys cys cys thr AA TGC TGC ACA	phe TTC
30	thr ACA	88	asp	val phe GTC TTC	thr ACT	asn AAT	5 S	met ATG	476 cys TCC	th ACA
	val GTG	265 cys TGT	glu asn GAA AAT	val GTC	thr	gla CAG	val GTA	arg AGA	1 ys	₹8
	leu 17	11e ATC	£8	asp	ale SA	pro CCT	1ys	ala lys GCA AAA	thr	ala CCT
	ser lys leu val thr asp ICC AAG TTA GTG ACA GAT	tyr IAT	val	ala lys asp GCA AAG GAT	tyr IXI	glu glu pro gln asn leu 11e 1ys GAA GAC CCT CAG AAT TTA ATC AAA	thr lys lys val pro gln val ser ACC AAG AAA GTA CCC CAA GTG TCA	448 glu ala lys arg met pro cys ala GAA GCA AAA AGA ATG CCC TGT GCA	val GTC	asn
25	ser ICC	261 ala iys tyr ile cys glu asn gln GCC AAG TAT ATC TGT GAA AAT CAA	g g SA	321 glu ala lys asp val phe leu gly GAG GCA AAG GAT GTC TTG GGC	351 Iys thr tyr glu thr thr leu glu lys AAG ACA TAT GAA ACC ACT CTA GAG AAG	381 val glu glu GTG GAA GAG	thr Acc	£8 88	471 asp arg val thr lys cys cys thr GAC AGA GTC ACC AAA TGC TGC ACA	510 giu phe asn ala giu thr phe thr phe his ala asp ile cys thr leu CAG TIT AAT GCI GAA ACA ITC ACC TIC CAT GCA GAT ATA IGC ACA CTT
35	231 val GTT	261 ala GCC	291 ala GCC	321 glu GAG	351 Iys AAG	381 val GTG	411 tyr TAC	441 pro	471 asp GAC	91 u 6AG

	(1790)	(1883)	(2002)
_	550 pro lys ala thr lys glu gln leu lys ala val met asp asp phe ala ala phe val glu lys cys cys lys ccc AAG GCA ACA AAA GAG CAA CTG AAA GCT GTT ATG GAT GAT TTC GCT GCT TTT GTA GAG AAG TGC TGF AAG	567 570 cos phe ala glu qlu qly lys lys leu val ala ala ser gln ala ala leu qly leu ter ter tec cac cac cac cac cac cac cac cac cac c	(GAAAGAAAATGAAGATCAAAAGCTTATTCATCTGTTTTTTTT
5	558 559 cys cys TCC TCC	CATE	ATT
	ys c AG 1	ATCA	ATA.
)to 1	۶۹	VAAC
	yal ç STA (3 5	TAA
10	e P	7 58 1 7	etata
	ala CCT] S	נעכני
	ala GCT	a1a 600	CCAA
	phe TTC	ala GCT	WW6
	asp CAT	580 n 19 CA	CTCT
15	dsp CAT	AGT) (110
	met ATG	818 CCA	111
	val GT	ala cct	TTCT
	ala GCT	va)	GTTT
20	- 1,4) Jeu	ATCT
	A CTC	\$ 3	IATE
	£ 3	× × ×	AGCT
	2 9 2 9	5 S	CAAA
٥٢	0 r 4 ½	6 6	AGAT
25	S&O A Chr	570 a glu c cAG	ATCA
	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	2 S	SA S
	2 S 2 S	567 cys pt TCC TI	§
	ā 0 8 5	7 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	rer TAG
30	15 A	Ju ti	ter ter NTGAGAATAAC
	. ¥ 3	× 9	te CATG
	'al 1	asp asp lys glu thr CAC CAC CAC ACC .	CTAC
	leu v	SAC C	וכאפל
35	531 glu leu val 1ys his lys GAG CTC GTG AAA CAC AAG	561 ala 6 CCT 0	ter ter Catctcagcctaccatgagaataaga

TCATTTTGCCTCTTTTCTCTGTGCTTCAATTAATAAAATGGAAAGAATCTAA.... 20AA (2078)

Following are examples which illustrate procedures, including the best mode; for practicing the invention. These examples should not be construed as limiting. All percentages are by weight and all solvent mixture proportions are by volume unless otherwise noted.

5 Example 1 Isolation of Messenger RNA

Human liver mRNA was obtained following the procedure of Chirgwin, et al [Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294-5299]. Immunoprecipitation of albumin containing polysomes was performed according to Taylor and Tse [Taylor, J.M. and Tse, T.P.H. (1976) J. Biol. Chem. 251, 7461-7467]. In vitro translation of mRNA was carried out in a reticulocyte cell-free system, following the instruction of the manufacturer (New England Nuclear). The translation products were separated electrophoretically according to Laemmli [Laemmli, J.K. (1970) Nature 227, 680-685.

Cloning Procedures Example 2

15

30

Double stranded cDNA was synthesized as described previously [Law, S., Tamaoki, T., Kreuzaler, F. and Dugaiczyk, A. (1980) Gene 10, 53-61]. It was annealed to PstI-linearized pBR322 DNA [Rolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crossa, J.H. and Falkow, S. (1977) Gene 2, 95-113 that had been tailed with 15 dG residues/3'-terminus [Dugaiczyk, A., Robberson, D.L. and Ullrich, A. (1980) Biochemistry 19, 5869-5873]. The annealed DNA was used to transform E. coli strain RR1, as detailed previously [Law, 25 S., et al., <a>Ibid.]. The albumin clones were selected using the colony hybridization method of Grunstein and Hogness [Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965], with [32p]-labeled cDNA synthesized with the immunoprecipitated polysomal mRNA as template.

As shown in Example 5, plasmids pHA36 and pHA206 were deposited in E. coli HB101 hosts. The plasmids were obtained from E. coli RR1 hosts, described in this example, and transformed into E. coli HR101 by standard procedures well known to those of ordinary skill in this The E. coli RR1 hosts were lysed and then centrifuged to 35 separate the chromosomal DNA, cell DNA and plasmid DNA. The plasmid DNA, remaining in the supernatant, is precipitated with ethanol and the precipitate is resuspended in buffer, e.g., TCM (10mM Tris·HCl, pH 8.0, 10 mM CaCl₂, 10 mM MgCl₂). The cells for transformation are prepared as follows: 120 ml of L-broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl) are inoculated with an 18 hour culture of HR101 NRRL B-11371 and grown to an optical density of 0.6 at 600 nm. Cells are washed in cold 100 mM NaCl and resuspended for 15 minutes in 20 ml chilled 50 mM CaCl₂. Bacteria are then concentrated to one-tenth of this volume in CaCl₂ and mixed 2:1 (v:v) with annealed plasmid DNA, prepared as described above. After chilling the cell-DNA mixture for 15 minutes, it is heat shocked at 42°C for 2 minutes, then allowed to equilibrate at room temperature for ten minutes before addition of L-broth 10 times the volume of the cell-DNA suspension. Transformed cells are incubated in broth at 37°C for one hour before inoculating selective media (L-agar plus 10 µg/ml tetracycline) with 200 µl/plate. Plates are incubated at 37°C for 48 hours to allow the growth of transformants.

15 Example 3 Mapping of Restriction Endonuclease Sites

Restriction endonucleases were obtained from Rethesda Research Laboratories and New England Biolabs and were used according to the manufacturers' instructions. The digested DNA fragments were analyzed electrophoretically on agarose [Helling, R.B., Goodman, H.M. and Boyer, H.W. (1974) J. Virol. 14, 1235-1244] or acrylamide [Dingman, C., Fisher, M.P. and Kakefuda, T. (1972) <u>Biochemistry</u> 11, 1242-1250] gels.

Example 4 DNA Sequencing

35

phosphatase (Worthington) and labeled at the 5'-ends with polynucleotide kinase (Boehringer-Mannheim) and Y[32p]ATP. Following digestion with a second restriction endonuclease and electrophoretic separation of the fragments, DNA sequence determination was done according to the procedure of Maxam and Gilbert [Maxam, A. and Gilbert, W. (1980) Methods Enzym. 65, 499-560] and the degradation products were separated electrophoretically on 0.4 mm acrylamide gels as described by Sanger and Coulson [Sanger, F. and Coulson, R. (1978) FEBS Letters 87, 107-110].

Example 5 Recombinant Plasmids pHA36 and pHA206

As disclosed in Example 2, albumin clones were selected by hybridizing to the enriched albumin cDNA probe. Plasmid pHA36 contained the largest insert of an albumin cDNA sequence. Both plasmids pHA36 and pHA206 have been deposited in a viable <u>E. coli</u> host in the

permanent collection of the Northern Regional Research Laboratory (NRRL), U.S. Department of Agriculture, Peoria, Illinois, U.S.A. Their accession numbers in this repository are as follows:

HB101(pHA36) - NRRL B-12551

HB101(pHA206) - NRRL B-12550

5

 $\underline{\text{E. coli}}$ HB101 is a known and widely available host microbe. Its NRRL accession number is NRRL B-11371.

NRRL B-12550 and NRRL B-12551 are available to the public. upon the grant of a patent. It should be understood that the availability of these deposits does not constitute a license to practice the subject invention in derogation of patent rights granted with the subject instrument by governmental action.

E. coli RR1 and E. coli HB101 are known and widely available host microbes. Their NRRL accession numbers are NRRL B-12186 and NRRL B-11371, respectively.

pBR322 is a well known and widely available plasmid. It can be obtained from the following host deposit by standard procedures:

NRRL B-12014 - E. coli RR1 (pBR322).

YEp6 is a well known and widely available yeast episomal plasmid.

20 It can be obtained from the following host deposit by standard procedures:

E. coli HB101 (YEp6) - NRRL B-12093.

Example 6 Assembly of the Serum Albumin Gene

Assembling the pieces together is a straighforward task of restriction enzymology. There is only one MspI site in the overlapping
DNA sequence of the two cDNA clones. Two enzymatic steps of (i) MspI
digestion of the two DNAs, followed by (ii) the use of ligase, an
enzyme that seals DNA fragments, will give the desired product.
Although two other undesired DNA species will also be obtained in the
course of this recombination reaction, both of them will differ substantially in size. Thus, separation and isolation of the desired DNA
species will be achieved.

The assembled DNA clone can be used to transform two types of cells:

(a) Escherichia coli

- (b) Saccharomyces cerevisiae
- (a) The vector of choice is plasmid pRR322, the same that has

been successfully used for cloning of the two fragmented pieces of the serum albumin cDNA.

(b) In order to transform yeast with the serum albumin structural gene sequence, the DNA must be inserted into one of the existing yeast plasmid vectors. This can be accomplished by taking advantage of the fact that several restriction endonuclease recognition sequences are absent from the cloned serum albumin DNA. Synthetic EcoR1 DNA linkers can be ligated to the DNA fragment containing the serum albumin sequence followed by insertion (ligation) into one of the yeast plasmid vectors, e.g., YEp6, at the Eco R1 cloning site. The fused chimeric plasmid can be used to transform yeast according to an established procedure [Hinnen, A., Hicks, J.B. and Fink, G.R. (1978) Proc. Natl. Acad. Sci. USA, 75, 1929]. YEp6 can be obtained from the NRRL repository, as disclosed supra.

15 Example 7 Expression of the Serum Albumin Gene

The main body of the structural gene will be transcribed by the E. coli or yeast enzymes. If little or no albumin is produced with the selected host, then an Escherichia coli promoter DNA sequence carrying an initiation codon, i.e., ATG, can be ligated at the begin-20 ning of the serum albumin structural gene. Such elements are known and available, e.g., lac promoter used for the expression of human interferon gene in E. coli [Proc. Natl. Acad. Sci. 77, 5230 (1980)]; source of promoter DNA [Proc. Natl. Acad. Sci. 76, 760 (1979)]. Also, see Nature, Vol. 281, October 18, 1979. It has already been 25 documented that such Escherichia coli promoter sequences function well in the expression of foreign genes in Escherichia coli [Mercereau-Puijalon, O., Royal, A., Cami, B., Garapin, A., Krust, A., Gannon, I. and Kourilsky, P. (1978) Nature 275, 505; and Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Grea, R., Hirose, 30 T., Kraszewski, A., Itakura, K., and Riggs, A. (1979) Natl. Acad. Sci. USA 76, 106]. For expression in yeast, see Rose, M., Casadaban, M.J. and Botstein, D. (1981) Proc. Natl. Acad. Sci. USA 78, 2460 and 4466. Example 8 Screening of Clones Producing Albumin

Immunological methods can be used to detect small amounts of albumin made in a bacterium. Flat disks of flexible polyvinyl are coated with the IgG fraction from an immune serum and the disks are pressed onto an agar plate so that antigen released from an in situlysed microbial colony can bind to the fixed antibody. The plastic

disk is then incubated with the same total IgG fraction labeled with radioactive iodine so that other determinants on the bound antigen can in turn bind the iodinated antibody. Radioactive areas on the disk expose X-ray film during autoradiography and thus identify colonies producing the protein which is being screened for. Detailed protocols of this procedure have been published [Broome, S. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA, 75, 2746]. The purification of human serum albumin can be accomplished by using procedures well known in the art. For example, procedures disclosed in a chapter by T. Peters: Purification and Properties of Serum Albumin, in: The Plasma Proteins, Putnam, Ed. Academic Press, New York, 1975, can be used.

The work described herein was all done in conformity with physical and biological containment requirements specified in the NIH Guidelines.

CLAIMS

- 1. Plasmid pHA36, having a restriction endonuclease pattern as shown in the drawing.
- 5
 2. Plasmid pHA206, having a restriction endonuclease pattern as shown in the drawing.
- 3. E. coli HB101 (pHA36) having the deposit accession number $_{10}\,$ NRRL B-12551.
 - 4. $\underline{\text{E. coli}}$ HB101 (pHA206) having the deposit accession number NRRL B-12550.
- 5. A microorganism modified to contain a nucleotide sequence coding for the amino acid sequence of human serum albumin; said nucleotide sequence is as follows:

20

25

	(30)	170)	(092	(350)	(044)	(330)	(029)	(710)	. 6
	AGC	20 lys AAA (170)	50 ala GCA (260)	8 7 E	110 pro ccA	140 try TAT C	170 aln CAA (620)	200 cys TGT (230 glu GAA (:)
		phe ly		thr le	asu p	leu t	169 1 cys 9 TG: G		230 ala qlu GCA GAA (300)
5	leu ohe CTC TTT	asn pl	alu phe CAA TTT	ala ti GCA AG		tyr le	168 16 cys cy TGT TG	leu lys CTC AAG	
					6 %		•	₽ ₹	alu phe CAG TTT
	bhe TTT	alu alu GAA GAA	val thr GTA ACT	thr val	s asb A CAT	s lys	4 6 P	ACA C	a alu r cAc
	-10 1eu CTT	₽ \$	val GTA		25	leu lys TTG AAA	phe thr TTT ACA	CAG	ala OCT
	7	leu aly TTG GGA	₹ 8	55 57 1	0 PI 3	phe leu lys lys TTT TTG AM AM	af E	ele lys qin arq GCC AAA CAG AGA	pro lys CCC AAA
10		19 1	AT	3 5	leu gin his lys asp asp TTG CAA CAC AAA GAT GAC	thr phe ACA TTT	ala CCT		£ 55
		lys asp leu aly AM GAT TTG GGA	leu val asn glu TTA GTG AAT GAA	75 asp lys leu cys GAC AAA TTA TOC	phe leu gin his lys TTC TTG CAA CAC AAA	glu glu thr GAA GAG ACA	tyr lys ala TAT AAA GCT	ser	gin arg phe pro CAG AGA TTT CCC
		phe lys TTT AAA	Je TA	asp GAC	phe TTC	6 g	<u>x</u>	ser TCG	91 AGA
		phe TTT	40 val lys GTA AAA	, 5	101 673	£ \$	tyr TAT	ala GCT	g Ja CAG
- -		10 his arg cat con	40 his val cat gta	phe TTT	등 등 옷	130 asn AAT	160 arg Acc	190 178 AG	220 leu ser CTG AGC
15		his	his CAT	le CTT	AAT	GAC GAC	1,3 8	919 555	Je CTG
		ala GCT	S T	thr leu ACC CTT	2 5	F. 54	a ta	₽	679
			<u>₽</u> X	his thr CAT ACC	gly arg asn GGC AGA AAT	phe TTT	phe TTT	190 asp qlu qly lys ala ser GAT GAA GGG AAG GCT TCG	ala trp ala val ala arg CCA TGG CCA GTA GCT CGC
		glu val CAG GTT	ag E	leu CTT			phe TTC		val GTA
20			o V	ser leu TCA CTT	glu pro GAA CCT	thr	le CTT	3 5	ala val CCA GTA
20		lys ser AAG AGT	¥ 55	173 AA	£ X		leu CTC	glu leu arg GAA CTT CGG	5 5
		CAC	34 gin eys CAG TGT	asp lys GAC AAA	lys gln AAA CAA	124 met cys ATG TGC	glu leu GAA CTC	asp GAT	ala trp CCA TGC
		2 2 C	gln CAG	62 cys TGT	8 5 8 7 8 7 8 7	val		leu CTC	<u>\$</u> \$
		1 asp ala his GAT GCA CAC	Jeu g	AAT T	91 Cys 2	asp GAT	ala pro	1ys	phe lys
25		- F 53	30 tyr TAT	976	8 55 75 75	120 val GTT	150 tyr TAT	180 CA	210 818
		arg :	gln 1	ala GCT	asp GAC	glu r	phe	150	۳. کې د ت
		0 <u>9</u> 는	sla c	ser a	ala a	2 33 2 33	tyr TAC 1	Ten CTC	ag 8
			phe 111	g ng g					
		p gly val GGT GTG	ala p	asp g	glu met CAA ATG	arg leu val arg CGA TTG GTG AGA	his pro	177 ala cys GCC TGC	1 g
30		9-0 9-0 9-0 9-0		ala a	gly g GGT C	1eu v 176 9	2 Q	12 G	\$ X
			leu ile TTG ATT	val a	tyr g TAT G	87.9 J	* < ₽5	lys ala	- 4
		-1 tyr ser TAT TC	val le			≅ Ω 8 સ	141 glu ile ala arg arg GAA ATT GCC AGA AGA	4.5	ser leu gln lys phe gly AGT CTC CAA AAA TTT GCA
		ala tyr GCT TAT	21 ala leu val GCC TTG GTG		glu thr GAA ACC	111 asn leu pro AAC CTC CCC (를 등 는 등	ala asp GCT GAT	# 5 # #
		. 5 2 5	a le	1 s thr A ACA		1 C C1	7 T	ઇ ਛ ન • ~	
35	•	36.T	21 818 800	2	81 arg CCT	111 asn AAC	₹ 5 8	171 als CCT	201 8.13 600

	(890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
5	260 1eu CTT	290 11e ATT	320 ala GCT	350 ala GCC	380 1eu CCT	410 8rg CCT	# # 60 P i s	470 ser AGT (500 175 AAA (
	ala asp ccc cAC	289 cys TGC	tyr TAT	Je C	bro CCT		AAA AAA	Val.	, erg	530 leu val CTT GTT
	နှုန် ဂင္ဂ	h is	AAC	2 A	1 × × ×	1 e	438 cys TGT	5 Y	val p	a)a 1 GCA C
=	AGG	ser TCC		leu CTG	glu phe lys GAA TTF AAA	Jeu CTG	437 cys TCT		tyr v TAC 0	7 5 C1 0
	55 55	1 xs	316 9ys 1GC	Jeu CTG	25.	ale OCC	173 AA 9	lys t	رک بر ان بر	¥
10	88p GAT	ag CA	val GTT	val leu leu leu GTG CTG CTG CTG	asp glu CAT CAA	asn ala Jeu leu val AAT GCG CTG TTA GTT	AGC	qlu lys thr GAG AAA ACG	asp qlu thr GAT GAA ACA	520 ser qlu lys glu arq qln lle lys lys gln thr ala TCT GAG AAG AGA CAA ATC AAG AAA CAA ACT GCA
	\$ 18 GCT		asp val GAT GTT	val	ala lys val phe asp GCC AAA GTG TTC GAT	e la CAG	430 leu qly lys val qly ser CTA GGA AAA GTG GGC AGC		Sp q	78 A
		leu CTG	Lys AAG	val	val	phe TTC	vai GTG	leu his TTG CAT	val a	1e 3
	253 glu cys GAA TGT	pro CCT	glu ser lys GAA AGT AAG	3er TCT	17s		173 AA	va) 1	o Pro	- X
	250 leu leu CTG CTT	2 ₹	£ \$5	ty. TAC	ala SCC	tyr lys TAC AAA	91y	461 cys v TCT (leu g CTG O	£ 2
15	250 1eu CTG	280 91u 1ys CAA AAA	310 val GTT	340 asp CAT	370 tyr TAT	400 glu tyr lys GAG TAC AAA	430 Jeu c	460 461 leu cys TTA TGT	490 ala leu glu val asp GCT CTG GAA GTC GAT ⁻	520 glu a
	gly asp GGA GAT	279 973 TGT	phe TTT	pro CCT	369 157 157	4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	AAC	gln 1 CAG 1	ser a	% % %
		278 573 100	88p GAT	his CAT	중동	3 5	\$ 59 \$ 59	asn o	phe 1	520 ser qlu lys glu TCT GAG AAG GAG
	N I S	5 ₹	als GCT	arg arg AGA AGG	his CAT	al n	ser arg TCA AGA	leu asn CTG AAC	15°5	ž 5
20	245 246 oys cys his TGC TGC CAT	leu lys glu CTG AG GAA	ala		pro CCT	GAG GAG	val	val GTC	8. 433 7. 433	3 5
	245 033 1GC	le CTG	7 E	ء اء دره	ass CAT	phe TTT	pro thr leu val glu val CCA ACT CTT GTA GAG GTC	tyr leu ser val val TAT CTA TCC GTG GTC	క్ట్రీ ప్ర	514 Ile cys thr leu ATA TGC ACA CTT
	£ ₹	17.8 AA	ser TCA	tyr	ala CCA	leu CTT	val	3er 700	arg AGG	514 6ys 1GC /
	thr Acc	AGT	pro	g)u	ala CCT	glu GAG	leu	leu CTA	asn arg AAC AGG	lle o
	h is	ser TCC	leu TTG	tyr IAT		392 cys TGT	th ACT	7.7		asp CAT
25	val GTC	11e ATC	asp GAC	Jeu TTG	361 0ys TGT	390 gin esn CAA AAT	8 Y 2		480 ser leu val TCC TTG GTG	510 his ala asp CAT GCA GAT
	240 178 AAA	270 ser TCG	300 ala CCT	330 phe TTT	360 375 155	330 g1n CAA	420 thr	450 91u GAA	480 TCC	510 his ala CAT GCA
	thr	asp GAT	pro CCT	gly met GGC ATG	1ys AAG	1 ys	ICA	818 SCA	g Ju	Phe TTC
	lea CTT	₽Ş	met ATG	917	glu lys	gin asn leu ile CAG AAT TTA ATC	gln val	448 oys ala TGT GCA	ACA	thr Acc
30	asp CAT	glu asn GAA AAT	9 6 6	Jeu 77	led CTA	leu TTA	gla CAA	ဥ္သည	477 cys TGC	phe thr TTC ACC
	th ACA	g GA	asp CAT	phe TTC	thr thr ACC ACT	AAT	. 5 g	met ATG	476 cys TCC	SC th
	481	265 933 1GT	asn AAT	val	thr	gln	val GTA	arg met AGA ATG		gh.
	ser lys leu val thr asp TCC AAG TTA GTG ACA GAT	tyr lle TAT ATC	val glu asn c	asp GAT	2g 6	glu pro	1 ys	13s	thr ACC	ala SCT S
	Lys	tyr TAT	val GTG	1ys AAG	tyr TAT	g lu GAG	1ys AAG	ela lys GCA AAA	val GTC	MT (
35	36. 700	lys AG	28€.	321 glu ala lys asp val phe GAG GCA AAG GAT GTC TTC	thr	ole S¥3	th Acc	ale GA	arg val thr lys AGA GTC ACC AAA	ohe i
	231 val GTT	261 ala GCC	291 ala GCC	321 91u CAC	351 173 AAG	381 val GTG	411 tyr TAC	441 Pro CCT	471 asp CAC	501 glu phe asn ala glu thr phe GAG TIT AAT GCT GAA ACA TTC
	O K	~ ទី	2	888	# # \$	38	¥4 8	\$ E S	47 asi	55 P

	_	_	_
	(1790)	570 phe ala glu glu qly lys lys leu val ala ala ser gln ala ala leu qly leu ter TTT GCC GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GGC TTA TAA CATCACATTTAAAAG (1883)	AAGAAAATGAAGATCAAAAGCTTATTCATCTGTTTTTCTTTTTCGTTGGTGTAAAGCCAACACCCTGTCTAAAAACATAAATTTCTTTAA (2002)
	560 173 MG	Ç K	₹
_	559 560 cys lys TGC AAG	ter TAX	E
5	558 578 160	ACAT	MIT
-	1ys AAG	CATC	CATA
	a) u	T ter	XX
	540 559 559 560 1ys ala val met asp asp phe ala phe val glu lys cys cys lys AAG GCA ACA AAA GAG CAA GCT GTT ATG GAT TTC GCT CCT TTT GTA GAG AAG TGC TAG AAG	Jeu TTA	CTAA
10	a E	414	CTGT
	ala act	1 E	ΩS
	5 5	200 CCC	ξÇ
	a be	ela GCT	AAAG
	550 4 55 54	580 gln CAA	2010
15	\$ 5 7	ser AGT	SCT TO
	met ATG	ء ۽ در	II
	42 CTT	ala GCT	1101
	# F3	vai GTT	STTF
20	.	CTT	ATCT
	Zeu CTG	!ys X X	ATTC
	e a	1 × ×	œ11
	a ja	aly ccr	¥
	_ ₹	gle GG	Q GTC
25	S&O thr	570 glu GAG	3
	# Y	313 CC	¥
			¥€
	2 22	567 693 160	700
30	1ys	glu thr GAG ACC	ter ter
	2 2	9 Ju 50	ter TÇAÇ
	1 , ₹	asp lys	V
	leu val lys his lys CTC GTG AAA CAC AAG	asp GAT	ter ter ATCTCAGCCTACCATGAGAATAAGAA
	Lea	es CA CA	CTCA
35	531 91u GAG	561 ala GCT	2

TCATTTTGCCTCTTTTCTCTGTGCTTCAATTAATAAAAATGGAAAGAATCTAA..... 20AA (2078)

6. Nucleotide sequence of the cDNA of human serum albumin, said nucleotide sequence is as follows:

5	(170)	(260)	(350)	(440)	(330)	(620)	(710)	230 91u GAA (300)
	20 173 AAA	S = S	98 CTT	110 Pro CCA	140 try TAT	170 Q1n CAA	200 0y3 TGT (230 910 GAA (
	phe TTC	Phe	thr	AAC	1 eu	169 cys TGC 0		
	AAT	ala CA	8 20	asb GAC	tyr TAC 1	168 0ys 1GT	leu lys CTC AAG	phe ala TTT CCA
10		the ACT	731 CTT	asb GAT	lys AA	g Pa	25	
10	glu glu GA GA	val GTA	ACA	1ys	lys lys AAA AAA	thr ACA	aln ara CAG AGA	ala qiu phe GCT GAG TTT
		16. AS		P.I.S.	15	phe TTT		
	leu aly TTG CGA	AAT	75 leu cys TTA TCC	gln his CAA CAC	phe TT	ala (ala lys GCC AAA	pro lys CCC AAA
	asp GAT		lys AA	15 T	ACA T	ala GCT	ser .	
15	phe lys asp leu qly TTT AAA GAT TTG GGA		asp GAC	phe TC	glu thr GAG ACA	lys AAA	3er 700	220 Ser gln arg phe AGC CAG AGA TTT
	Phe TT	40 val lys leu GTA AAA TTA	£ 55 € 45		₹ 8	tyr TAT	ala GCT	ala CAG
	to his arg phe CAT CGC TTT	40 481 GTA	2 pg 11.1	100 101 qlu eys GAA TGC	130 asn qlu AAT GAA	160 AGG	190 173 AAG	220 3er AGC
	his	CA S	leu CTT	asa AAT		\$ ₹	A 500	Jeu CTG
00	818 CCT	CA T	th Acc	2	phe his asp TTT CAT GAC	ata CCT	alu aly GAA GGG	ala arg leu GCT CGC CTG
20	val GTT	2 ₹	his CAT	58	ala phe his GCT TTT CAT	ohe TTT	asp GAT	ala GCT
	916	pro phe CCA TTT	ser leu TCA CTT	pro		glu leu leu phe GAA CTC CTT TTC	550 550	vai GTA
	ser AGT			g y GA	thr	Jeu	leu CTT	e te
	1ys AAG	34 gin oys CAG TGT	asp lys GAC AAA	g y	124 cys TGC	leu CTC	g &¥	t.79
25	h1s CAC		asp GAC	1 × 8		og y	asp	e te CCA
		leu gln CTT CAG	62 93 1GT	# X	asp val GAT GTG	ala pro ccc ccc	leu CTC	173 AA
	asp GAT		AAT	91 0ys 1GT			1ys AG	phe TTC
		30 gln tyr CAG TAT	8 g 8	90 cys TGC	120 glu val GAG GTT	150 tyr 1AT	8 c c c c c c c c c c c c c c c c c c c	210 818 GCT
			ala GCT	asp		phe TTT	leu. TTG	AC A
30		ala CCT	15 A	ala ect	£ 55	tyr IAC	leu CTG	95. GA
		phe TT	95 95	glu met GAA ATG	arg leu val arg	his pro CAT CCT	17 293 100	phe gly TTT CCA
		a1a CC	esp CAT	g Y	val GTG	5.5 CA T	\$18 CCC	phe TTT
		11e ATT	ala CCT	gly ccr	3 5	ala arg arg CCC AGA AGA	lys ala	gin lys CAA AAA
35		leu 110	val GTT	tyr IAT	55	\$ 5 \$	lys AA	g S¥
		21 ala leu val leu lle ala phe GCC TTG GTG TTG ATT GCC TTT	63. 643 161	glu thr GAA ACC	97° CCC	818 CC	asp GAT	lei CTC
		150 170	thr ACA		leu	141 glu 11e GAA ATT	ala CCT	ser AGT
		21 ala CCC	2	81 arg CCT	111 asn AAC	2 m 2 m 3 m 3 m 3 m 3 m 3 m 3 m 3 m 3 m	171 ala CCT	201 ala GCC

5	(890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	. (1520)	(1610)	520 glu arg gin fie lys lys gin thr ala leu val GAG AGA CAA ATC AAG AAA CAA ACT GCA CTT GTT (1700)
	260 1eu CTT	290 11e ATT	320 ala GCT	350 ala GCC	380 leu CCT	\$ 10 8 10 CGT	440 hIs CAT (470 ser AGT (Sno 1ys AAA (530 val CIT (
	asp GAC	289 633 160	tyr	3 E	57	val a	1 ys 4	val s	Pro 13	2 2 2 2
	ala CCG	h is	AAC	25	lys r	leu v	438 cys 1 TGT A	910 CCA G	val p	2 5 2 5
10	ACC ACC	Ser Tor	173	leu CTG	phe lys TTT AAA	Jeu 1 CTG 1	437 4 cys 0 TGT 1	thr p	tyr v.	# 5 # 1:
	aso GAC	<u>₹</u>			2 K	818 000	lys c	lys t	thr ty ACA T	∓ ¥
	asp CAT	E SY	316 val cys GTT TGC	val leu leu GTG CTG CTG	asp alu	gln asn ala leu leu CAG AAT CCG CTG TTA	ser 1 AGC A	ole SAC A	g)u ti	£ ₹
	al a	15 15			ake 4	מנפּ א	91.4 CCC A	his q CAT G	asb GAT G	£ 5 ₹ ₹
	253 cys TGT	5 CT	lya asb AAG CAT	val val GTC GTG	val phe GTG TTC	phe q TTC C	val q ctc c	11G C	val a: GTC Q	= 5 € 5
15	કે કે	Pro CCT	ser AGT		1 × ×	× 1	lys val AAA GTG	val 1. GTG T	glu v	= = = =
	lei CTT	173 AA	35 64	tyr	ala lys val phe GCC AAA GTG TTC	tyr 1ys TAC AM	93,4 1	461 cys v TGT G	leu g CTG G	6 S
	250 1eu CTG	280 q lu CAA	310 Val	340 pro asp tyr ser CCT CAT TAC TCT	370 tyr 8 TAT 0	400 glu t GAG 7	430 leu qly CTA CGA	460 461 leu cys val leu his qlu lys thr TTA IGT GTG TTG CAT GAG AAA ACG	490 ala le GCT C	520 glu au GAG AC
	asp GAT	279 cys TGT	phe TTT	Pro CCT	369 cys TGC	7 6.5 7 6.5	asn 1 AAC C	gin 1 CAG T		6.9.50
20	99		asp CAT				arg a	asn g AAC C	phe ser TTT TCA	leu ser glu lys CTT TCT GAG AAG
20	5 E	£ %	ala CCT	arg his AGG CAT	his alu CAT CAA	aln leu CAG CTT	ser a	leu asn CTG AAC	cys p	ser q TCT G
	246 cys TGC	lys AG	ala CCT	arg AGA	pro his CCT CAT	SAG O		val l	Pro c	leu s
	245 glu cys CAA TGC	leu lys CTG AAG	ser leu ala TCA TTA GCT	ala arg GCA AGA	asp CAT	phe TT	glu val GAG GTC	val v GTG 0	5 Y	
	ele GA	₹	3er 7CA	tyr	818 CCA		val g	3er v TCC G	arg a	514 cys thr TGC ACA
25	ACG	ser lys AGT AAA	6 5	ale SA	ala	glu leu GAG CTT	leu v CTT O		asn a	Z = 2
	val his GTC CAC	ser TCC	leu TTG	7.7	2 1 a c c c c c c c c c c c c c c c c c c	392 cys (TGT (thr leu val ACT CTT GTA	tyr leu TAT CTA	val a	asp Ile CAT ATA
	val GTC	11e ATC	8 5 SC 5	leu TTG	361 cys TGT	AAT 1	5 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	asp t GAC T		
	240 173 AAA	270 ser TCG	300 818 CCT	330 Phe 111	360 cys TCC	3%0 91n CAA	\$20 thr p	450 gtu a GAA G	480 ser leu TCC TTG	510 his ala CAT GCA
	thr	asp CAT	pro	met ATG	1ys	¥	ser (818 g CCA C	glu s	phe h
30	Jea CTT	gla CA	met ATG	91 80 91 91	91u GAG	11e J	val s	448 0ys a TGT G	thr g ACA G	thr pl
	asp CAT	asn AAT	g GG				o Hg		73 73 74 75 74	
	thr ACA	265 cys glu asn TGT GAA AAT	asp	phe leu TTC TTG	thr leu ACT CTA	asn leu AAT :TTA	pro gin	et p	476 477 cys cys TGC TGC	בי בי קיד
	val CTG	26 5 cy3 TGT		val	thr	45	val p	చ్చి	lys c AA T	3 ₹
35	Jec 17	11e ATC	glu asn GAA AAT	asp	glu thr GAA ACC	pro gln CCT CÁG	lys v	* ₹	val the 1ys GTC ACC AAA	12 g
33	lys AG	tyr 11e IAT ATC		1ys		75 B	lys 1 AAG A	- Y	val ti GTC A	. S
	231 val ser lys leu val thr asp GTT ICC AAG TTA GTG ACA GAI	1ys	glu val CAA CTC	ala lys CCA AAG	thr tyr ACA TAT	glu glu pro gln asn leu GAA GAG CCT CAG AAT TTA	the 1	glu ala lys arg met pro GAA GCA AAA AGA ATG CCC	arg va AGA GI	ie a
	231 val GTT	261 ala lys GCC AAG	291 ala ccc	321 91u a 6A6 0	351 lys thr AAG ACA	381 val 9 GTG 0	411 tyr thr lys lys val TAC ACC AAG AAA GTA	##1 pro 9 ccr G	471 asp ai CAC A(501 glu phe asn ala glu thr phe CAG TTF AAT GCT GAA ACA TTC
							4 4 F	4 50	4 00	v ⊕ G

5	(1790)	570 phe ala glu glu qly lys lys leu val ala ala ser gln ala ala leu qly leu ter TTT GCC GAG GAC GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GCC TTA TAA CATCACATTTAAAAG (1883)	NAGAAAATGAAGATCAAAAGCTTATTCATCTGTTTTCTTTTTCGTTGGTGTAAAGCCAACACCCTGTCTAAAAACATAAATTTTTTTAA (2002)	
		ر کر	₹	
	559 5 0ys 1 1GC /	ter ITAA	ICTT	
	558 5 eys 9 TCC 1	(CAT	AATT	
	1 × 8 × 1	CATC	CATA	
10	260	Tèr T≷	AAA	
	CTA	1 ted	CTAA	
	3 E	58	CTCT	
	ele GCT	14 114	CACC	
15	550 Iys ala thr Iys glu gin leu lys ala val met asp asp phe ala phe val glu lys cys cys lys AAG GCA ACA AAA GAG CAA GCT GTT ATG GAT GAT TTC GCT CCT TTT GTA GAG AAG TGC AGG		7)	
	Phe	ala GCT	TA.	
	550 885 CAT	580 91 CA	cctc	6)
	es GAT	ser AGT	CGTT	(207
00	met ATG	\$ \frac{1}{2}	1111	\$
20	val GTT	8 F	TTCT	:
	sls GCT	va1	ICTT	20
	\$ \$	100	. ATCI	
	Lec CTO	¥	ratte	CTAA
25	£ 3	,	AGCT	GAAŢ
	650	4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	CAA.	Ğ
	\$ \$ \$	# §	AGA T	AATG
	Ş ∄ Ş	570 a 91u c 646	ATCA	₩
	3 3	1 a a a	₹	TWI
30		s ph	¥	CAAT
	£ 0	7 5 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	# §	CCTT
	\$ Y	t d	ter ter ITGAGAATAA(cra
	2 Z	5 G	ter ATGA	TTC
35		6 P ₹	TACC	rcrı
	6 K	8 S	Yecc	71000
	531 glu leu val lys his lys pro GAG CTC GTG AAA CAC AAG CCC	561 ala asp asp lys glu thr cys GCT GAC GAT AAG GAG ACC TGC	ter ter CATCTCAGCCTACCATGAGAATAAGAGA	TCATTITGCCTCTTTTCTCTGTGCTTCAATTAAAAAATGGAAAGAATCTAA 20AA (2078)
	2 2 2	8 8 %	5	ĭ

7. Nucleotide sequence coding for the prepeptide of human serum albumin, said nucleotide sequence is as follows:

	•	
	(30)	
5	- C	
	leu ohe	
	1ec CTC	
	95 P P P P P P P P P P P P P P P P P P P	
10	2 I E	
10	-10 CTT C	
	-10 -11e ser leu bhe	
	11e ATT	
	9he 111	
15	r chu	-
	p r o trp val tlu phe E	
	י ני	
	7 17 17 17 18 18 19 19 19 19 19 19 19 19 19 19 19 19 19	
20	- Hei	
	GCAC!	
	-18 Met GCTTTCTGTGTGAACCCCACAGCCCTTTGGCACA ATG	
	ומכנכ	
	CAC	
25	MCCC	
	1670	_ 25
		2 5
	110	r o -1 I phe arg arg G TTT CGT CGA
30	861	7 2 2
		2 \ 2 \ 2 \
		-1 -6 p ser ala tyr ser arg gly val TCG GCT TAT TCC AGG GGT GTG
		- 9 S
		tyr a
35		201
		301 201

8. Nucleotide sequence coding for pro human serum albumin, said nucleotide sequence is as follows:

5	(071)	50 818 GCA (260)	(350)	(440)	(330)	(029)	(017)	(300)
	8	818 878	80 CTT	110 Pro CCA	140 try 1AT	170 41n CAA	200 cys TGT	230 910 CAA
	phe TTC	phe TTT	thr	AAC	1 Per	169 673 160	NG AG	8 8 CCA
	asn	g y	618 GCA	3 S C C C C C C C C C C C C C C C C C C	tyr TAC	168 0ys 1GT	leu CTC	9 te
••	ole GAA	thr	val GTT	asp	<u>*</u>	ale GA	2 A	d of c
10	4 P		thr	1ys		ACA		ala GCT
	55	glu val GAA GTA	27. 100.	his CAC	leu lys TTG AAA	a Pe	lya aln AAA CAG	Ay Ay
	Jeu 11G		75 leu oys TTA TCC	al CAA	phe 111	ala GCT	313	200
	esp CAT	val GTG	A S	16e	ACA L	ala ala GCT GCT	ser TCT	ag E
15		40 his val lys leu val asn CAT GTA AAA TTA GTG AAT	SS D	Phe TTC	26		100 100	97.9 ACA .
	phe Jys TTT AAA	40 glu asp hfs val 1ys 1eu GAA CAT CAT GTA AAA TTA	916 CGA	10 10 10 10 10	alu alu GAA GAG	tyr lys TAT AAA	ela GCT	gin
		40 CTA	phe TT	8 4 5 8 4 5	130 asn AAT	160 AGC	130 AC 53	220 36T AGC
	10 his arg CAT CGG	EAT CAT	3 5	asn AAT	asp GAC	***	9. 200	leu CTG
	# J &	S S	Acc	2 Y		als CCT	₹ 8	D SS
20	val	₹ ĕ	leu his thr CTT CAT ACC	419	ala phe his GCT TTT CAT	glu leu leu phe phe ala lys CAA CTC CTT TTC TTT GCT AAA	asp	e la CCT
	95 646	phe TTT	leu his CTT CAT	or CCT	ala GCT	phe TTC	5 00	
-	lys ser AG AGT	e Y	ser	of A	thr ala ACT GCT	leu phe CTT TTC	leu CTT	ala val CCA GTA
	lys AG	34 0ys TGT	₹ ¥	E AS	124 137 137	leu CTC	glu leu GAA CTT	
25	ele his GCA CAC	gln gln CAG CAG	asp	<u>₹</u> ¥	ATG ATG	5 €	asp GAT	ala trp CCA TGG
23	919 CCA	gln gln CAG CAG	62 0ys 1GT	ala GCA	val	200	leu CTC	1ys
	asp GAT	leu CTT	asn AAT	91 093 TGT	asp	818 CC	1ys AAG	phe TTC
	- 58	30 tyr TAT	8 5 8 8 8	90 87.50 17.00	120 val CTT	150 tyr 1AT	180 CC 5	210 ala GCT
	o phe arg TTT CGT	91n CAG		asp GAC	a) u	phe TTT	leu 11G	8 5 8 7 8
30	. # E	ela CCT	ser ala TCA GCT	ala CCT	క్షి స్ట	7. 2	leu CTG	g P
	, ** 1 CTC	phe TTT	alu GAG	met ATG	Ag Ag	873 CC1	17. 8% 100	91. 86.
	-6 pro rgglyvalphean ccccccccttc	ala ccc	asp	glu met ala GAA ATG GCT	leu val arg p TTG GTG AGA C	arg his pro I	1 sts	phe TTT
	9-19 ACC	11e ATT	ala GCT	gly	leu TTG	5 A	ala	1ys AA
		leu 11G	val	tyr	ş Ş	\$ 5 5	13s	ol b
35		21 ala leu val leu lle ala phe GCC TTG GTG TTG ATT GCC TTT (53 035 TGT	thr	or o	ala arg GCC AGA	177 asp lys ala ala cys CAT AAA GCT GCC TGC	leu gin lys phe CTC CAA AAA TTT
		leu 776	ACA th	£ ₹	leu CTC	11e ATT	ala GCT	AGT
		21 818 000	12 5. 15. S. A.	81 879 CCT	111 asn AAC	141 glu lle GAA ATT	171 ala CCT	201 ala :

5	(890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
	260 1eu CTT	290 11e ATT	320 818 CCT	350 ala GCC	380 CCT	410 810 CGT	640 FAT	470 ser AGT	500 173 AAA (530 val GTT (
•	S Sp	289 cys TGC	tyr TAT	Je CTT	pro CCT		¥	val 9	pro 1	Jea v CTT G
	# 58 88	N I S	asn AAC	£ 5	1ys	leu val TTA GTT	638 Cys TCT	0 433 CCA C		ala 1 GCA C
10	AG A	lys ser AAA TCC	173 AA	leu leu leu arg CTG CTG CTG AGA	phe TTT			thr Acc c	tyr v 14C G	thr a
	\$ 5 5 5	17s	316 6y3 TGC	5 S	ole GAA) 920 818	437 ser lys cys AGC AAA IGT	lys t	thr t	4 4 CA 4
	ese GAT	alu GA	val	leu CTC	S S S	AAT (AGC /	alu Jys GAG AAA	alu t	8.¥
	e le GCT	leu leu CTG TTG	asp GAT	val	phe TTC	ala CAG	4 to 000		asp q	8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8
	253 cys TGT	leu CTG	glu ser lys asp val GAA AGT AAG GAT GTT	val GTC	val	lys phe gin asn ala leu AAA TTC CAG AAT GCG CTG	val qly GTG GCC	leu hís TTG CAT	val a	ain lie lys lys ain thr CAA ATC AAG AAA CAA ACT
15	, 28 X	Pro CCT	ser AGT	ser TCT	7 A A	lys phe AAA TTC	1 × × ×	val I	91c V	aln 1 CAA A
	250 leu leu CTG CTT	178		tyr	# F		45 58		leu g CTG C	arg g AGA C
	250 1eu CTG	280 41u 6&4	310 val GTT	340 asp CAT	370 tyr TAT	400 91u tyr GAG TAC	436 leu qiy lys val CTA GGA AAA GTG	460 461 Jeu cys TTA TGT	490 Ser ala leu glu val asp qlu thr tyr val TCA GCT CTG GAA GTC GAT GAA ACA TAC GTT	520 glu arq GAG AGA
	839 CAT	279 cys TGT	ag I	pro CCT	369 C.Y.S			gln J	ser a	
20	his gly CAT GGA	278 078 1GC	asp CAT	MIS CAT	글	leu aly CTT GGA	arg asn AGA AAC	AAC O	phe s	glu 1ys GAG AAG
20		16 8	ala	ACC	his CAT	gin leu gly CAG CTT GGA	ser TCA	tyr leu ser val val leu asn TAT CTA TCC GTG GTC CTG AAC	5. 10.	ser g
	245 246 glu cys cys GAA TGC TGC	leu 1ys CTG AAG	leu ala TTA CCT	ACA ACA	52	\$ 55 \$4		val GTC	67 67 7	
	240 245 lys val his thr glu oys AAA GTC CAC ACG GAA TGC	leu CTG	ser leu TCA TTA	8 18 GCA	asp CAT	phe TTT	glu val GAG GTC	va1 GTG	2 43 20 43	514 Ile cys thr leu ATA TGC ACA CTT
	₹ 3	ser lys AGT AAA	pro ser CCT TCA	glu tyr GAA TAT	818 CCA	glu leu GAG CTT	val GTA	100 100	arg a	514 cys t TCC A
25	t the		Pro CCT	916 68	ala CCT	alu GAG	2 H	Jeu CTA	asn a	514 Ile cys ATA TCC
	240 lys val his thr AAA GTC CAC ACG	270 ser ile ser TCG ATC TCC	Je 0	tyr IAT	ala SCC	392 cys TCT	thr	7.4		asp 1
. .	×21	11e ATC	asp GAC	1ec 776	361 cys TGT	asn	pro CCA		leu val TTG GTG	ala ccA
•	240 178	270 8er TCG	300 ala ccT	330 phe TTT	360 973 160	3% gln C&	420 thr ACT	450 glu asp GAA CAC	480 ser 1	510 h18 d CAT 0
_	A th	asp	Pro CCT	gly met GGC ATG	glu lys GAG AAG	<u>1</u> 33	ser TCA		glu s	phe h
30	Je CTT	£ \$	met ATG		glu GAG	11e ATC	val GTG	448 cys ala TCT CCA	thr 9	thr p
	2 gs	glu asn CAA AAT	asp glu GAT GAG	Jeu TTG	Jed CTA		gln CAA	500	973 (age 1
•	ser lys leu val thr asp TCC AAG TTA GTG ACA GAT	33	83p CAT	phe TTC	tyr glu thr thr leu TAT GAA ACC ACT CTA	asn AAT	. 67 900	met ATG	476 477 cys cys TCC TCC	4 2
	× 41	265 11e oys ATC TGT	asn AT	val GTC	thr	gla CAG	val GTA	E S	<u>\$</u> \$	35 A
35	150	11e ATC	36	asp CAT	₹3	pro CCT	1 xs	¥ ¥ 8	ihr 1	ola g
	¥	tyr IAT	val GTG	lys AG		o de c	lys AG	818 CCA	/al (ISN 8
		261 ale lys tyr GCC AAG TAT	giu val glu asn a GAA GTG GAA AAT C	321 glu ala lys asp val phe leu GAG GCA AAG GAT GTC TTG	351 1ys thr AAG ACA	g]r CA	411 tyr thr lys lys val pro TAC ACC AAG AAA GTA CCC	2 8 8 8	5 73 0 73	she a
	231 val CTT	261 a18 GCC .	291 818 GCC	321 g l u GAG	351 133 AAG	381 val CTG	411 tyr TAC	441 pro glu ala lys arg met CCT GAA GCA AAA AGA ATG	471 asp arg val thr lya cya cys CAC AGA GTC ACC AAA TGC TGC	501 glu phe asn ala glu thr phe GAG TTT AAT GCT GAA ACA TTC
	•									

531 531 540 551 550 550 550 550 550 550 551 650 550 550 550 551 650 550 550 550 550 550 550 550 550 550	5	(1790)	(1883)	(2002)	
531 540 550 551 551 552 553 553 553 554 555 555 554 555 555 555		560 178 AAG	¥œ	14	
531 540 539 531 540 550 558 558 558 559 559 550 558 550 558 550 558 550 558 550 558 550 558 550 558 550 558 550 558 550 558 550 558 550 558 550 558 550 558 550 558 550 558 550 550		559 579 TGC	ter TTAA	1011	
531 530 531 530 530 530 530 530 530 530 530 530 540 550 550 550 551 551 551 551 551 551 55		558 678 TGC	ACAT	AATT	
531 530 531 530 531 540 540 540 540 540 540 550 551 551 551 551 551 551 551 551 55		1ys AAG	CATC	CATA	
531 540 531 540 540 541 540 540 541 540 540 541 540 540 541 540 541 540 541 540 541 540 541 540 541 540 541 541 541 541 541 541 541 541 541 541	10	n n cyc	ter TA	¥	
531 540 530 531 540 540 540 540 540 540 540 540 540 540		val GTA	leu 17A	ICT.	
531 531 540 530 531 540 540 540 540 540 540 540 540 540 540		phe TTT	5 33	is Tel	
531 540 531 540 531 540 540 550 550 551 561 567 567 570 580 581 580 580 580 580 580 580 580 580 580 580		818 GCT	leu TTA	VCACC	
531 540 531 540 531 540 540 550 550 551 561 567 567 567 561 567 560 560 560 560 560 560 560 560 560 560	15	ala GCT	ala ငက	%)	
531 530 531 540 550 550 550 550 561 567 567 567 567 567 567 568 567 567 567 568 567 568 567 568 568 567 568 568 569 561 569 569 561 560 560 560 560 560 560 560 560 560 560		phe	ala CCT	Ϋ́	
531 540 540 540 540 540 540 541 540 541 540 541 541 541 542 541 542 543 543 544 540 544 540 541 541 541 541 541 541 541 541 541 541		550 asp CAT	580 414 CA	SCTG	6
531 540 531 540 531 540 540 540 540 540 540 540 540 540 540		asp CAT	AGT	CCTT	(207
531 S40 531 GC CTC CTC AAA CAC AAA AAA GAG CAA CTC AAA GCT CTT S61 ala asp asp lys glu thr cys phe ala glu glu qly lys lys leu val ala CCT CAC CAT AAG CAC TCC TTT GCC GAG GGT AAA AAA CTT GTT GCT CATCTCAGCCTACCATGAAAAAAAAAAAAAAAAAA		met ATG	a1 a CCA	TITI	¥
531 540 531 540 540 540 551 540 552 553 554 557 557 551 557 551 557 550 551 557 550 551 557 550 551 557 550 551 557 550 551 552 551 557 550 550 550 550 550 550 550 550 550	20	val GTT	ala CCT	TTCT	
531 S40		els cct	val	CTTT	2
531 S40 S40 GAG CTC CTG AAA CAC AGG CCC AAG GCA ACA AAA GAG CAA CTG S61 ala asp asp lys glu thr cys phe ala glu glu qly lys lys GCT GAC GAT AAG GAG ACC TGC TTT GCC GAG GCT AAA AAA CTT CAC CAT AAG GAG ACC TGC TTT GCC GAG GCT AAA AAA CATCTCAGCCTACCATGAGAAAAAGAAAATGAAAGATTATTGTAAA		1,48	CTT	ATCT	
531 S40 S40 glu leu val lys his lys pro lys ala thr lys glu gln GAG CTC GTG AAA CAC AGG CCC AAG GCA ACA AAA GAG CAA 561 ala asp asp lys glu thr cys phe ala glu glu qly lys GCT GAC GAT AAG GAG ACC TGC TTT GCC GAG GGT AAA GCT CACCTACCATGAGAAAAAAAGAAAAGAATCAAAAGCTT		leu CTG	¥	ATTO	TAA
531 S40 glu leu val lys his lys pro lys ala thr lys glu GAG CTC GTG AAA CAC AAG GCA ACA AAA GAG 561 ala asp asp lys glu thr cys phe ala glu glu qly GCT GAC GAT AAG GAG ACC TGC TTT GCC GAG GGT CATCTCAGCCTACCATGAGAAAAAAAGAAAAAAAAAA	25	₽ 2	\$ \$	1201	71 4 4
531 S40 glu leu val lys his lys pro lys ala thr lys cAG CTC GTG AAA CAC AGG CCC AAG GCA ACA AAA 561 ala asp asp lys glu thr cys phe ala glu glu GCT GAC CAT AG GAG ACC TGC TTT GCC GAG CAG CATCTCAGCCTACCATAGAGAAAGAAAGAAGAATG	20	g g	45	¥	7444.
531 GAG CTC CTG AAA CAC AAG CCC AAG GCA ACA 561 ala asp asp lys glu thr cys phe ala glu GCT GAC GAT AAG GAG ACC TGC TTT GCC GAG CATCTCAGCCTACCATGAGAAAATGAA		¥.	88	GATC	77.
531 GAG CTC GTG AAA CAC AAG CCC AAG GCA 561 ala asp asp lys glu thr cys phe ala GCT GAC GAT AAG GAG ACC TGC TTT GCC ter ter CATCTCAGCCTACCATGAGAAAGAAAA		Seo thr	576 916 676	35	7
531 GAG CTC GTG AAA CAC AAG CCC AAG 561 ala asp asp lys glu thr cys phe GCT GAC GAT AAG GAG ACC TGC TTT CCT CACCCTACCATGAGAAAA		ala ccA	2 3 3 4 5 C C C C C C C C C C C C C C C C C C	**	
531 glu leu val lys his lys pro GAG CTC GTG AAA CAC AAG CCC 561 ala asp asp lys glu thr cys GCT GAC GAT AAG GAG ACC TGA CATCTCAGCCTACCATGAGAATAAGA	30		a E	NA NA	
531 glu leu val lys his lys cAG CTC GTG AAA CAC AAC 561 ala asp asp lys glu thi GCT GAC GAT AAG GAG ACC		pro CCC		¥ \$	
531 glu leu val lys hie GAG CTC GTG AAA CAC 561 ala asp asp lys glu GCT GAC GAT AAG GAC		17.	ACC th	**************************************	Ì
531 glu leu val lys GAG CTC GTG AM 561 ala asp asp lys GCT GAC GAT AM		hta CAC	. g	ter VTGA	
531 glu leu val GAG CTC GTG 561 ala asp asf GCT GAC GAI	25	17.	, ly	ZYCC	
531 glu leu GAG CTC 561 ala ast GCT GAC	35		S 35	VGCC1	
91 6 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5				rctc/	
		53. 50.00	, 9 K	รี	

TCATTITIGECTCTTTTCTCTGTGCTTCAATTAATAAAAATGGAAAGAATCTAA..... 20AA (2078)

9. Nucleotide sequence coding for the pre pro human serum albumin, said nucleotide sequence is as follows:

5	(30)	(170)	(1560)	(350)	(pat)	(330)	(620)	(710)	(300)
	AGC	20 178 AA	8	Se CTT	110 Pro	140 TAT	170 qln CAA (200 cys TCT (230 916 644
		phe TTC			AAC	leu TX	169 Cys TCC		230 Iys ala alu phe ala alu AAA GCT GAG TTT GCA GAA
	leu ohe CTC TTT	AAT	alu phe CAA TTT	ala thr GCA ACT	ase GAC	tyr TAC	168 cys 1GT	leu] CTC /	ع <u>ب</u> ا
10	-10 leu leu phe leu CTT CTT TTT CTC	a]r CA	thr Act	val	ase CAT	lys AA	of S SA	lys ain ara leu lys AAA CAG AGA CTC AAG	lys ala qlu phe AAA GCT GAG TTT
10	1 E	2 ¥3		thr ACA	1ys	lys AAA	AÇ thr	aln d	2 ts
	-12 CTT	aly alu alu GGA GAA GAA	₹\$	2 % 2	his lys CAC AAA	15 T	phe II	1ys	% %
	367	1ec 110		75 asp lys leu cys thr GAC AAA ITA TGC ACA	CA S	qlu qlu thr phe leu lys GAA GAG ACA TTT TTG AAA	ala CCT	818 CC C	200
	116 ATT	6 ₹ 5		₹	leu aln TTG CAA	ACA T	ala GCT	3er 2	ag E
15	ag E	₹	leu val TTA GTG	\$ 5 5	phe TTC	2 2	£ \$	3er :	5 2
	Acc thu	g be	lys AAA	4 2	5 % D	alu GAA	TAT /	ala s	ole /
	p r o lys trp val tlu phe AAG TGG GTA ACC TTT	5 to 000	40 val lys GTA AAA	2 Phe 11 TT	100 101 qlu oys phe GAA TGC TTC	130 asn AAT	160 arg tyr lys ACG TAT AAA	190 lys ala ser AG GCT TCG	220 leu ser gin arg CTG AGC CAG AGA
	t 1	₹ ₹	CAT CAT	3 5	AAT	asp GAC		750	35 CTG
00	-16 Het lys ATG AAG	#1# CCT	asp CAT	SG th	2 2	his CAT	ala lys GCT AAA	£ §	ala arg leu GCT CGC CTG
20	Fet ATC	glu vel GAG GTT	₹ §	\$ F	91. CC 2	phe TTT	ohe TT	S. C. T.	val ala arg GTA GCT CGC
		a) u	phe TTT	CT E	Pro CCT	ala phe CCT TIT	age 1	glu leu arg asp GA CTT CGC GAT	
		lys ser AAG AGT	£ 5	15 P	of S		5 15	leu CTT	ele SCA
		ly.	4 8 P	<u>\$</u>	gin glu CAA GAA	124 135	leu CTC	3	trp ala TGC GCA
25		FI S	gln CAG	ase CAC	2× ₹	met ATG		asp CAT	lys ala trp ala val AAA GCA TGG GCA GTA
		* 5	gla CAG	62 cys TGT	s s CCA	val GTG	pro glu	leu asp CTC GAT	lys ala AAA GCA
		- 5 5 5 1	Je CTT			asp GAT	ala pro OCC CCG	1ys AG	
		- 53	30 tyr TAT	60 glu asn CAA AAT	90 91 cys cys TGC TGT	120 val CTT	150 tyr TAT	180 pro 1ys CCA AAG	210 ala phe GCT TTC
		phe arg	ele CAG	ala GCT	# 5 CAC	pro glu	phe TTT	1ec 11C	9.75 V3.
30		3 g =	ala	7 S	als CCT		tyr TAC	leu leu CTG TTG	glu arg GAA AGA
		p r gly val GGT GTG	ala phe GCC TTT	9 Se 3					917
		aly ccT	als OCC	es of TA	glu met CAA ATG	val	his pro	် ရေး ပိပ္ပ	a be
		-1 -6 ser arg g TCC AGG (11e ATT	ata CCT	aly GGT	160 170	\$ 54 \$	ala CCT	\$. \$.
35		- # 5 5	116	E TE	174	E S	2 2	173 AA	£ 5
JJ		ele tyr GCT TAT	val CTC	53 Cys TCT	th ACC	111 asn leu pro arg leu val arg AAC CTC CCC CCA TTG GTG AGA	<u>و</u> 2	asp lys ala ala oys GAT AAA GCT GCC TGC	ser leu gln lys phe AGT CTC CAA AAA TTT
		5 to	12	ځ ک <u>ې</u>	glu thr GAA ACC	leu CTC	11e	ala GCT	AGT
		36T	21 als	2 × 5 €	819 CCT	111 850 AC	141 glu fle GAA ATT	171 ala GCT	201 ala GCC

				~	=	=	=	=	5	=	=
5		(890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
	260	S E	290 11e ATT	320 ala CCT	350 818 GCC	380 1ee CCT	\$10 810 CGT	440 hls CAT	470 ser AGT	Sno 1ys AAA	530 val GTT
		asp leu GAC CTT	289 eys 1GC	tyr TAT	leu CTT	oro CCT		\$ ₹		5 333	le CTT
		arg ala AGG GCG	CAC CAC	AAC	25 AGA	phe 1ys TTT AAA	1 es	437 438 cys cys TCT TCT	5 Y		818 CCA
10		AGG	ser his TCC CAC	1 XX	leu arg CTG AGA	phe TTT	leu leu CTG TTA	437 cys TGT	thr Acc	tyr TAC	thr
•		asp asp GAT GAC	lys AA	316 cys TGC	Jeu CTG	a Jr CAA	asn Bla AAT CCC	ser lys AGC AAA	1 A A A .	thr	£ \$
		esp GAT	₹3	316 val cys GTT TGC	ser val val leu leu leu arg TCT GTC GTG CTG CTG AGA	asp GAT	asn	ser	leu his qiu lys thr pro val TTG CAT GAG AAA ACG CCA GTA	₹ ₹	¥ }
		و ا	leu 11G	asp GAT	val	phe TTC	ela CAG	2 53	his		1ys AAG
	253	cys TGT	leu CTG	1ys AAG	val val GTC GTG	va) GTG	phe TTC	val qly GTG GGC	leu hís TTG CAT	val GTC	11e ATC
15		glu cys GAA TGT	pro CCT	ser	36. 101	ala lys val phe GCC AAA GTG TTC	lys phe gln AAA TTC CAG	1ys AA	val	# 6	£ \$
		leu CTT	\$ \$	glu ser lys asp GAA AGT AAG GAT	tyr TAC	ها درد	400 glu tyr GAG TAC	4) y	\$61 833 TGT	leu CTG	AGA 13
	250	le CTG	28 2 = 3 2 = 3	310 val GFT	340 asp GAT	370 tyr TAT	400 gly glu tyr lys phe gln asn ala leu leu val GCA GAG TAC AAA TTC CAG AAT GCG CTG TTA GTT	430 1eu CTA	7 FE	490 ala CCT	510 his ale asp lie oys thr leu ser glu lys glu arq qin lie lys lys qin thr ala leu CAT GCA GAT ATA TGC ACA CTT TGT GAG AAG GAG AGA CAA ATC AAG AAA CAA ACT GCA CTT
		S &	279 0ys 1GT	Phe TT	pro	369 cys 160	5 59		g]n CAG	Ser	X6
20		his gly CAT GGA	278 cys TCC	asp GAT	h1s CAT	GA 43	glu leu phe qlu gln leu GAG CTT TTT GAG CAG CTT	ser arg TCA AGA	tyr leu ser val val leu asn TAT CTA TCC GTG GTC CTG AAC	9 tr	56.92
20			5.83	ata CCT	arg AGG	A IS	es CAG	ser TCA	Jeu CTG	3 1 2 3	Ber TCT
	246	5 2 3	lys AG	ala GCT	ala arg GCA AGA	2 5	35	glu val GAG GTC	val	- £ 5	leu CTT
	245	glu oys oys CAA TCC TCC	ser lys leu lys glu AGT AAA CTG AAG GAA	pro ser leu ala ala CCT TCA TTA GCT CCT	al a GCA	ala ala asp. p GCT GCA GAT G	glu leu phe qlu gln leu GAG CTT TTT GAG CAG CTT	420 thr pro thr leu val glu val ACT CCA ACT CTT GTA GAG GTC	val GTG	క్ష్మి స్ర	514 asp 11e cys thr 1eu GAT ATA TCC ACA CTT
		골종	1ys A A	36. 104	glu tyr GAA TAT	* te CC	leu CTT	val GTA	ser TCC	arg AGG	514 0ys 100
25		his thr	AGT	pro CCT	g Jr GAA	ala CCT	a) ta GAG	leu	1eu CTA	asn AC	11e
		\$ 5 \$ 5	ser TCC) eu	tyr	ھڑھ 20	392 cys TGT	thr	tyr	leu val TTG GTG	asp GAT
		lys val	270 ser 11e TCG ATC	300 ala asp GCT GAC	leu TTG	361 cys TGT	390 gin asn CAA AAT	pro	450 glu asp GAA GAC	1eu 11G	al a
	240	\$	270 ser TCG	300 81a CCT	330 Phe TT	360 978 1GC	3% g1n CAA		450 glu GAA	480 3er 1CC	510 hts CAT
		thr	gin asp CAA GAT	met pro ATG CCT	gly met GGC ATG	glu lys GAG AAG	ile lys ATC AAA	ser TCA	448 cys ala TCT CCA	gle GA	thr phe ACC TTC
30		SE	g &	met ATG	917	ale GAG	ile lys ATC AAA	val	448 cys TCT	thr	thr
		\$ \f		26	Jeu TTG	led CTA	<u>3</u> ₹	gln CA	ည္ဆိ	477 cys TCC	phe TTC
		lys leu val thr asp AAG TTA GTG ACA GAT	glu asn GAA AAT	asp CAT	phe TTC	tyr glu thr thr TAT GAA ACC ACT	asn	. 57	met ATG	476 1ys oys AAA TGC	thr ACA
		val GTG	265 cys TGT	ase AAT	val GTC	thr Acc	gln	vel GTA	arg AcA	8.¥8	ag &
35		Jec 17	tyr lle TAT ATC	ale GA	asp GAT	ole SA	pro CCT	lys lys AAG AAA	<u>\$</u> ₹	thr	al a
33				glu val glu asn asp GAA GTG GAA AAT GAT	1ys AG	tyr IAT	a) a	1ys AG	8 50 CA	val thr lys GTC ACC AAA	asn AAT
		ser TCC	lys AG	glu GAA	321 glu ala lys asp val phe GAG GCA AAG GAT GTC TTC	351 1ys thr tyr glu thr thr AAG ACA TAT GAA ACC ACT	381 val glu glu pro GTG GAA GAG CCT	411 tyr thr lys lys val pro gln val ser TAC ACC AAG AAA GTA CCC CAA GTG TCA	441 pro glu ala lys arg met CCT GAA GCA AAA AGA ATG	چ 5 کا	phe TT
	231		261 ala CCC	291 ala GCC	321 91u GAG	351 1ys AAG	381 val GTG	411 tyr 1AC	441 pro	471 asp CAC	501 glu phe asn ala glu thr phe GAG TTT AAT GCT GAA ACA TTC

5		(1790)	580 gin ala ala leu qiy leu ter CAA GCT GCC TTA GCC TTA TAA CATCACATTTAAAAG (1883)	ter ter Catctcagcctaccatgagaataagaaaaaaaaaagatcaaagcttattcatctgtttttctttttggtgtaaggccagcctgtctaaaaagcataaatttgtttaa (2002)	
•			5	1 (2	
		178 AG	۲	₹ E	
		oys cys lys	a ₹	17.7	
	e V	8 y 3	ACA	MATI	
10		lys AG	CATC	CATA	
10		age Sy	5 ₹	Š	
		I V	3 ₹	₹	
		\$ =	≥ 8	TCTC	
			2 ₹	رزز	
15	•	2 L	2 2	¥C	
13		ن و ي و	® 5	¥	
	ş.	lys ala thr lye glu gin leu lys ala val met asp asp phe ala ala phe val glu lys cys cys lys AG GCA ACA ACA AM GAG CAA CTG AAA GCT GTT ATG GAT GAT TTC GCT GCT TTT GTA GAG GAG TGC TGC AAG	570 phe ala glu glu qly lys lys leu val ala ala ser gln ala TTT GCC GAG GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT	₹.	
	· •	. ĕ ∂	7 E D	5	78)
		* Q	A AG	rcer	(20
20		T AT	# B	E	₹.
20		8 L		110	
	,	2 S	<u> </u>	CTTO	20
•		₹ ₹) E	ATCT	
		3 5	₹ ₹	VTTC	₹
25		5 5	, ₹	CTT.	ATCI
		28	SG 2	¥	₹
		\$ ₹	g gr	Ą	ş
	7 4 5	Ş ţ	570 glu GAG	¥e	¥
		€ 53	2 16	WAT.	1
30		× ×	2 =	§	¥I.
				AGA	₹
		AG P	بر ج 2 ک	ter TAAG	755
		AC A	S C	ter ter TGAGAATAA	CTC
	•	glu leu val lys his lys pro GAG CTC GTG AAA CAC AAG CCC	561 ala asp asp lys glu thr cys CCT CAC CAT AAG CAG ACC TGC	te ATC	TCATTTTGCCTCTTTTCTGTGCTTCAATTAATAAAATGGAAAGAATGTAA 20AA (2078)
35		- 2 - 3	e ⊢ ∵ ≤	TACC	TCTI
		G <	ដូច	28	7000
	_		£ 33	ĊŢĊ	IIII
	<u> </u>	ร์ธั	8 2 2	5	5

- 10. A nucleotide sequence according to any of claims 6 to 9, in essentially pure form.
- 11. A DNA transfer vector comprising a nucleotide sequence as defined in claim 5.
- 5 12. A DNA transfer vector according to claim 11, transferred to and replicated in a micro-organism.
 - 13. A DNA transfer vector according to claim 12, which is a plasmid.
 - 14. A DNA transfer vector according to claim 13,
- 10 wherein the plasmid is pBR322 or YEp6.
 - 15. A process for preparing human serum albumin, which comprises culturing a micro-organism according to claim 5.
 - 16. A DNA transfer vector according to any of
- 15 claims 12 to 14, or a process according to claim 15, wherein the micro-organism is a bacterium or yeast.
 - 17. A vector or process according to claim 16, wherein the bacterium or yeast is <u>E. coli</u> or <u>Saccharomyces cerevisiae</u>.

1/1 6. 586 TAA Sst | 531/2 Restriction Endonuclease Map of Human Serum Albumin cDNA Clones 1.8 493/4 Tag l 382 419/0 450/1 Mbo II Hinc II Mbo II pHA36 Kilobases 325/6 Mbo II o; Hinf 1 269/0 œ Pst i (3611) Hpa II 182/3 Tag (8 ĸ. 4 pHA206 S7 Hinfl Hpa II (3658)